Advance Access publication on 6 May 2015
doi:10.1093/iwc/iwv012

Investigating Mental Workload Changes in a Long Duration Supervisory Control Task

Mark Boyer[1,2], M.L. Cummings[3], Lee B. Spence[2] and Erin T. Solovey[4,*]

[1] *MIT Humans and Automation Laboratory, Cambridge, MA, USA*
[2] *MIT Lincoln Laboratory, Lexington, MA, USA*
[3] *Humans and Autonomy Laboratory, Duke University, Durham, NC, USA*
[4] *Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA*
**Corresponding author: erin.solovey@drexel.edu*

With improving automation in many critical domains, operators will be expected to handle long periods of low task load while monitoring a system, and possibly responding to emergent situations. Monitoring the psychophysiological state of the operator during low task load may detect maladapted attention states in order to predict performance and facilitate a more effective workload transition during critical periods. This research explored the question of detecting anomalous attention states during transitions to high workload following extended periods of boredom using a non-invasive neuroimaging technique called functional near-infrared spectroscopy (fNIRS). Subjects at the point of lowest engagement and priming had a diminished hemodynamic response and performed worse on missile defense task, showing fNIRS may be useful for concurrent monitoring of the operator in such settings.

RESEARCH HIGHLIGHTS

- Functional near-infrared spectroscopy brain sensing is feasible for use in long duration (3 h) tasks.
- Hemodynamic response was diminished during the middle of a long duration, low task load simulation when engagement and priming were lowest.
- fNIRS did not detect a change in workload, but did reflect temporal changes in event onset, which could be used to automatically adapt a system when an operator is in a degraded attention state.

Keywords: human–computer interaction; user models; laboratory experiments; novel interaction paradigms; psychology; military; physiological adaptive computing; brain–computer interaction; mental workload; automation interaction

Received 19 June 2014; Revised 13 March 2015; Accepted 19 March 2015

1. INTRODUCTION

In a growing number of fields, the human has shed the role of direct controller and taken on the mantle of ‘system manager’ or ‘supervisory controller’. With improving automation, supervisory controllers face longer interludes between critical events, but those events are often more extreme and more demanding. Bainbridge (1983) articulated this paradox in the ‘irony of automation’, a concept that is readily applicable to many fields today. Missile defense is an extreme example of this: actual events are exceedingly rare, the stakes are incredibly high, and operators must act within seconds of an event beginning to properly address a threat. In order to perform at the highest levels of mental capacity during these critical events, operators should be properly attentive and engaged to the monitoring task so that they can quickly and competently respond. This research explores the processes involved in making a rapid transition from very low to very high mental workload in these low task load domains, in order to lay a foundation for psychophysiological adaptive automation in such supervisory control settings.

1.1. Workload transition

Mental workload has been one of the most widely studied fields of human factors research, but should be differentiated from task load. Task load is an objective measure of the actions required of an operator to execute a task, and is independent of experience or subjective response. Workload is the individual response to task load demands by an individual. For example, 2 air traffic controllers each managing 10 aircraft may report different levels of workload despite having the same taskload. Kahneman's Resource Theory proposes that human mental abilities can be modeled by a single 'bucket' of mental resources (Kahneman, 1973). When that bucket overflows, tasks must be shed and performance will suffer. Wickens (1984) expanded on this concept with Multiple Resource Theory by hypothesizing that there are different pools of resources based upon sensory or processing modality. The common thread is that humans are limited by their mental processing resources, which may be divided by function, and that the limits of these resources can be tested through experimentation.

Tasks often come in irregular intervals in arenas like defense, process control and aviation. In these critical fields, operators can experience extended periods of low task and workload followed by short periods of high task and workload, where they must overcome possible boredom, fatigue and distraction. Huey and Wickens (1993) cite factors such as uncertainty, surprise, task character and information processing as variables that can uniquely influence mental workload transition. This study examines the effects of surprise, uncertainty, task character and task timing on mental workload transition in the context of a missile defense simulation where high-workload events occur at an unknown and variable times (surprise) with an unknown number of targets to track (uncertainty) to discover the most important hemodynamic and behavioral features.

1.2. Functional near-infrared spectroscopy

Physiological measurement uses various signals of the body to measure how resources are being stressed and used (Kramer, 1991; Pattyn *et al.*, 2008; Veltman and Gaillard, 1996; Wilson, 2002). Psychophysiological measures of workload target the resources of the brain by measuring the elements of cognition using techniques such as electroencephalograph (EEG), magnetoencephalograph and functional magnetic resonance imaging (fMRI). Such techniques have been used in various settings to measure high and low workload (Berka *et al.*, 2005; Bunce *et al.*, 2011; Cui *et al.*, 2011; Dussault *et al.*, 2005; Hirshfield *et al.*, 2009; Sassaroli *et al.*, 2008; Warm *et al.*, 2009).

Functional near-infrared spectroscopy (fNIRS) (Fig. 1) is a relatively newer psychophysiological technique that uses the optical properties of hemoglobin to measure oxygen consumption near the outer surface of the brain. At certain wavelengths, near-infrared light passes through bone and tissue but is absorbed by the oxygen in the blood. By illuminating a portion of the brain and measuring the amount of light that is returned, it is possible to track oxygenated and deoxygenated hemoglobin levels over time (Chance *et al.*, 1998).

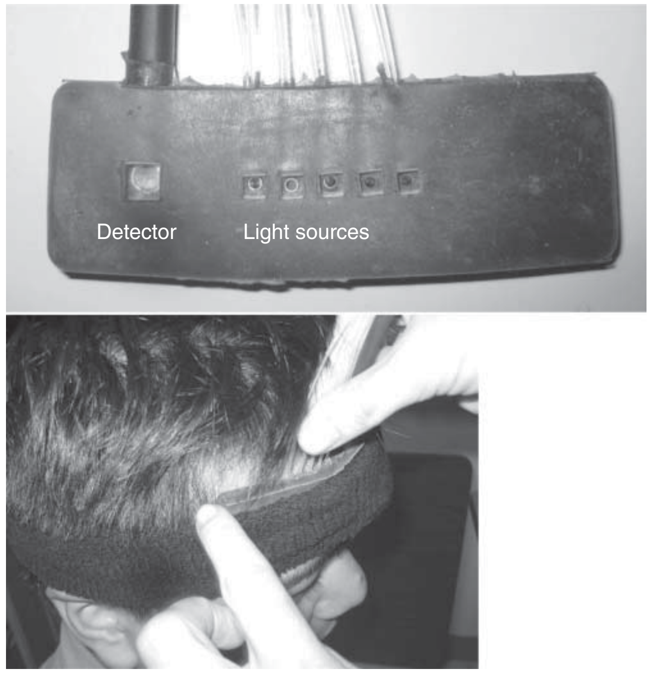

Figure 1. fNIRS sensor with linearly arranged light sources and detector (top). Two such sensors were placed next to each other on the forehead and secured with a headband (bottom) over the Fp1 and Fp2 locations.

Placement of fNIRS sensors on the forehead allows for probing the prefrontal cortex, which is situated behind the forehead area. Several studies have looked at the prefrontal cortex activity as a measure of mental workload in cognitive tasks (McCarthy *et al.*, 1994; Miller and Cohen, 2001; Scholkmann *et al.*, 2014; Tsujimoto *et al.*, 2004). With fNIRS measurements, a rise in levels of oxygenated hemoglobin and a decline in levels of deoxygenated hemoglobin have been reported in response to increased mental activity (León-Carrión and León-Domínguez, 2012). fNIRS can be used in a variety of settings to measure workload (Bunce *et al.*, 2011; Huppert *et al.*, 2006; Sassaroli *et al.*, 2008; Solovey *et al.*, 2012) and it shows promise for neuroergonomics (Derosière *et al.*, 2013) and sustained attention situations (De Joux *et al.*, 2013; Warm *et al.*, 2012). While lacking some spatial resolution compared with fMRI, several studies have demonstrated that fMRI and fNIRS measure similar responses (Cui *et al.*, 2011; Harrison *et al.*, 2013; Izzetoglu *et al.*, 2011; Schroeter *et al.*, 2006; Steinbrink *et al.*, 2006; Strangman *et al.*, 2002). From these findings, our hypothesis is that we would observe a similar response (increased oxy-hemoglobin and decreased deoxy-hemoglobin) in the fNIRS measurements during a rapid transition to a high workload period from a long duration, low workload period.

1.3. Workload measurement in supervisory control

fNIRS monitoring is attractive for measurement of workload in supervisory control domains because it is non-invasive and not as sensitive to movement as EEG. In addition, it is difficult to infer workload through operator interactions in low task load settings with little to no requirement for interaction until a low probability event occurs. Thus, fNIRS, as well as other psychophysiological devices, provide a continuous signal of operator state rather than discrete instances when the operator is physically interacting with the system through a computer or machine interface. These properties make physiological methods attractive for measuring operator state over time, particularly in low task load environments, in order to predict performance or vary automation level. However, because of the need to measure participants over long periods of time for such studies, physical comfort and low signal-to-noise signals become paramount, which is why fNIRS was considered the superior technology for this study.

This study aimed to measure the effects of extended periods of low task load on operator response during a critical event that causes workload to quickly rise. Previous work has not used fNIRS or any other psychophysiological device in such long-duration, low task load supervisory control settings (Schmorrow, 2005). In addition, while previous studies have used fNIRS to measure aspects of mental workload and modulate operator tasks based on fNIRS measurements (Afergan *et al.*, 2014; Ayaz *et al.*, 2012; Durantin *et al.*, 2014; Harrison *et al.*, 2013; Sassaroli *et al.*, 2008; Tsunashima and Yanagisawa, 2009; Wolf *et al.*, 2007), the experimental settings in these studies were necessarily artificial with low subject numbers and multiple events to achieve sufficient statistical power. These studies did not focus on low task loading, nor on a near-instantaneous dramatic change in workload. In this study, we chose to significanty increase the subject number (*a priori* estimated power of 0.8), but reduce the number of critical events to be more realistic for long duration supervisory control tasks. These details are discussed further in the next section, but ultimately this study aimed to assess how fNIRS responses differ in low and high task load environments, as well as investigate any associated performance degradations that could be associated with these changes.

2. EXPERIMENT

The experiment was based upon a notional ballistic missile defense mission using a desktop simulation. As noted earlier, this task is an example of a long duration, low workload task, where events are rare but critical and time sensitive. The participant acted as the sensor controller for three unmanned aerial vehicles (UAVs), each with a tracking sensor on board. The task of the participant was to allocate which UAV should track which incoming missile during a missile event.

The display (Fig. 2) was split across two monitors, with the primary display on the left and secondary display on the right. The primary display consisted of a Sensor Tracker Display window for each of the three sensors on the three

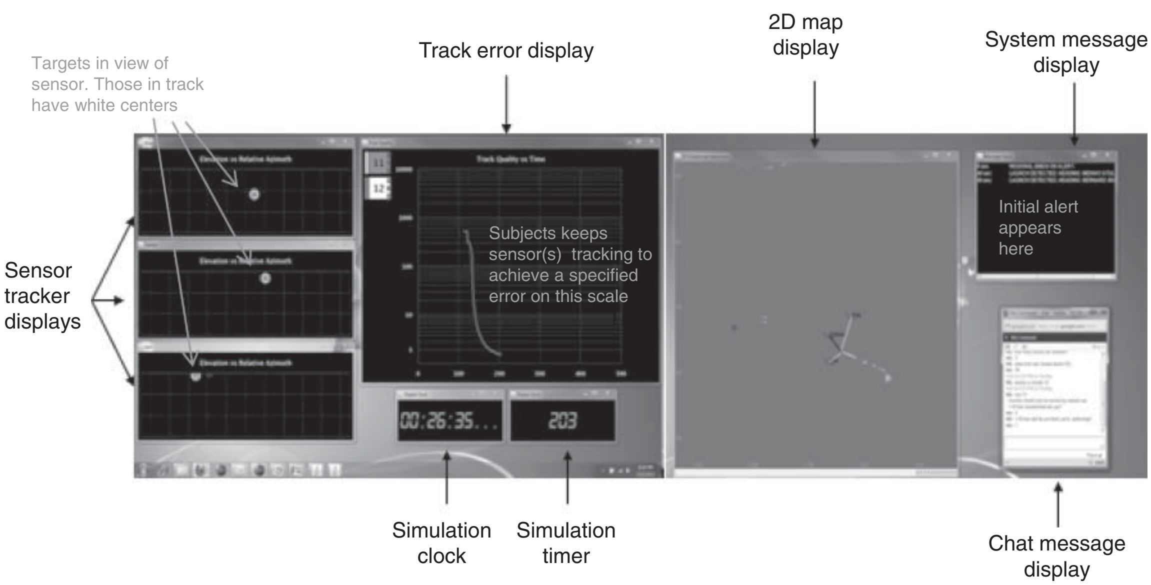

Figure 2. Missile defense simulation display. Display was spread over two screens, with the primary display on the left used for controlling the sensors and checking track error and the secondary display on the right for monitoring messages and the 2D map.

UAVs, and a Track Error Display window, where participants monitored their performance. The secondary display consisted of a 2D map, which gave a visual representation of which sensor was tracking which missile, a System Message Display which provides system status updates to the user, and a Chat Box, where users received scripted messages and answered questions at pseudo-random intervals.

In the simulation, a missile event consisted of a wave of missiles launched at times close together and lasted 100 s, which was unknown to the participants. When the event started, the operator received an alert on the System Message Display (Fig. 2, top right) that a launch has occurred. The target(s) that were in view from a sensor appeared in the respective Sensor Tracker Display box to the left of Fig. 2. At this point, the operator began to assign sensors to the target(s) with the goal of bringing the track error, displayed in the Track Error Display box down to a value that he/she had been given by the experimenter. They met this threshold by assigning a sensor to the target and waiting although there are options of using multiple sensors to speed the process. They were not told how long they would have, how many missiles they would have to track or when the missile event would occur. There were two waves that occurred in the experiment. We focus on the first wave, which occurred at 40, 100 or 160 min as described below in the Experiment Variables section. The second wave occurred at 180 min for all participants, but was not considered in this analysis of the effects of a mental workload transition from low taskload to high taskload for a novel event, since repeated waves are confounded by situational priming and learning effects.

When not actively engaged in a tracking task, participants monitored incoming chat messages on a separate screen and ensured the missile defense system was correctly operating. The chat messages had varying degrees of interaction from a personal question to a simple system status message. Precautions were taken in the implementation to prevent the secondary task from interfering with the primary task. During the low task load periods, chat box questions or statements were presented pseudo-randomly every 300–500 s. During the high task load period, questions were presented only every 15–20 s, and questions asked were simple to minimize time away from the primary interface. Although impossible to fully eliminate any possibility of interruption of the primary task, the location, frequency and salience of the chat box were tested and adjusted during pilot testing to ensure minimal distraction. Furthermore, the subjects were clearly instructed on the hierarchy of tasks at the beginning of the experiment and told to prioritize the mission over responding to chat messages.

Participants were trained on using the system with a 20-min self-paced slide tutorial and then given a simplified 5-min training mission to practice using the different features of the display. All participants were required to pass a knowledge check at the end of the training period. No participants required any remedial training. Subjects were paid $75 dollars to participate in the 3-h experiment and instructed that the top performer would also receive a $150 prize.

2.1. Experiment variables

The study was 3 × 2 between-subjects design. The first independent variable was *onset time* of the wave of missiles. Participants received the wave of missiles at either 40, 100 or 160 min, and the entire test session lasted 180 min. The second independent variable was *scenario difficulty*. Participants either received three missiles or six missiles during the incoming wave of missiles. Since there were only three tracking sensors to allocate, the six-missile scenario required a dynamic allocation of assets to achieve good performance, while the three-missile scenario allowed for a 1-to-1 allocation of sensors.

Because of the long duration nature of this task, the 'vigilance decrement' was an expected phenomenon, which is categorized by a decline in detection abilities often (but not always) occurring during the first 30 min of a vigilance task after which people reach a new equilibrium of diminished vigilance performance (Alves and Kelsey, 2010; Broadbent, 1958; Mackworth, 1948; Warm *et al*., 2008, 2009). The first onset time was set at 40 min, to occur after but close the expected 30 min threshold, which could also serve as a point of comparison with the later onset times.

2.2. Participants

Thirty subjects, ages 18–31, were recruited from a Northeast university to participate in this study. The average age was 21.3 years (s.d. 2.51) and the sample contained 12 males and 18 females, none of whom indicated any military experience. All participants were required to be right-handed, be a native English speaker, have normal vision and have no history of seizures, neurological disease or epilepsy. All participants completed a consent form. Further details about the gender and age within each of the six cells can be found in Table 1. There were no significant anomalies across the cells for individual differences for either age or gender.

Table 1. Gender and age within each of six blocks.

Difficulty	Onset time (min)	# Female	# Male	Average age
Easy	40	3	2	20.4
Easy	100	3	2	21.6
Easy	160	3	2	20.2
Hard	40	2	3	21.0
Hard	100	3	2	23.2
Hard	160	4	1	21.6

There were no significant anomalies across the cells for individual differences for either age or gender.

2.3. Data collection

Several forms of data were collected. The participants first completed a demographic survey that included age, gender, military experience, sleep history and video game experience. The participants also completed a Big Five personality index (NEO-FFI-3) (McCrae and Costa, 1999) and a Boredom Proneness Index (Farmer and Sundberg, 1986). After the experiment, the participants completed a NASA Task Load Index workload survey (TLX) (Hart and Staveland, 1988) and a custom survey with several questions to allow for feedback. Experiment data included two performance metrics from the missile tracking exercise (average final tracking error and percentage of missiles tracked better than the threshold), chat box message responses and chat box message response times. Video was collected for each participant for the length of the experiment, including a video of screen activity and video of the participant.

fNIRS data were collected using an ISS, Inc., Imagent device. This device recorded two wavelengths of light (690 and 830 nm) at 12 Hz for the entire experiment. Two probes were used, each containing four linearly spaced light sources and one detector (Fig. 1). The two probes were applied to each side of the forehead and secured with an elastic headband, with the sources centrally located above the prefrontal cortex (Fig. 1) mapping to the Fp1 and Fp2 locations of the 10–20 system for electrode placement of EEG. The source-detector distances were between 2.5 and 3.5 cm. Once the probes were secured to the participant's head, the software associated with the Imagent device was used to calibrate the sensors before beginning the simulation.

2.4. Data reduction

Once collected, the fNIRS data were processed to reduce each subject's response to a form suitable for statistical analysis using the Homer2 software package developed at Massachusetts General Hospital (Huppert *et al.*, 2009). First, the data were converted from raw light intensities to oxygenated hemoglobin concentration (HbO), deoxygenated hemoglobin concentration (HbR) and total hemoglobin concentration (HbT) levels in micromolar units. The data were then filtered using a 0.5 Hz low-pass filter to remove much of the noise and variability.

The period surrounding the event was isolated into a hemodynamic response function (HRF). The HRF included the 60 s before the start of the event as the reference and 100 s following the appearance of incoming missiles as the event, which was when the missiles disappeared from the operator's display. The expected hemodynamic response to increased workload involves a decrease in HbR with a rise in HbO and HbT. Therefore, the minimum HbR captures the relative greatest response and a negative percent change from the reference period indicates an increase in activity. Taking this into account, the maximum for each 100-s event period was found for each HbO and HbT signal, while the minimum was found for each HbR signal. We also took the average response over the same period for analysis, but did not expect it to be as informative as the local minimums and maximums. These event period maximums and minimums for HbO, HbR and HbT were then averaged across participants. This process was referred to as the average of maximum method and was also used for the reference period.

2.5. Statistical analyses

All data met normality and homoscedasticity assumptions (using the Kolmogorov–Smirnov and Levene's tests, respectively). The only data that did not pass these tests were the performance data (average final track error) and they were analyzed using Mann–Whitney and Wilcoxon tests.

3. RESULTS

3.1. Subjective workload

A one-way ANCOVA was conducted to investigate differences in NASA-TLX score related to difficulty level and event onset, controlling for age. The covariate, age, was marginally significantly related to the NASA-TLX score, $F(1, 23) = 3.40$, $P = 0.078$. There was also a significant effect of difficulty on NASA-TLX after controlling for the effect of age, $F(1, 23) = 4.6$, $P = 0.042$, so not surprisingly subjects sensed they were working harder in the more difficult scenario. Onset time was not significant.

3.2. Hemodynamic response

We examined the impact of missile wave onset time and missile wave difficulty on hemodynamic response. While we monitored the fNIRS signal throughout the course of the entire experiment, for this analysis, we examined the transition period from low to high. Thus, we were only interested in a specific time period, but we measured the operator continually through that specific time period. A reference hemodynamic state was measured by computing the average of maximum for the 60 s period before the event occurred. Since there were no indications to the participant of the impending event, this period captures the average state that may occur at any point in the time leading up to the event.

The average of maximum hemodynamic response across the four sensors for each participant after the critical event was then converted into a percent change from the reference. The mean HbO percent change from the reference to the event was 60.5% (s.d. 124.1%) while the mean HbR percent change was 81.5% (s.d. 150.4%). Since HbT is highly correlated with HbO, only the results from HbO are reported here. While not the focus of this analysis, we did also look at the average of the average hemodynamic response and saw similar trends

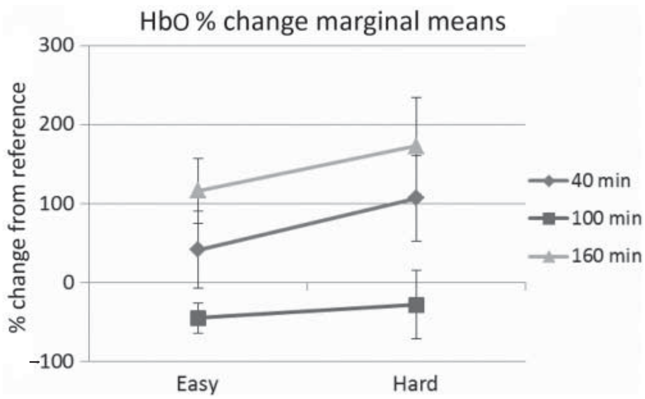

Figure 3. HbO % change shows 100-min onset was significantly lower than the 40- and 160-min cases, indicating a diminished response for participants during the middle of the experiment. No significant differences in HbO response for difficulty. Error bars indicate standard error.

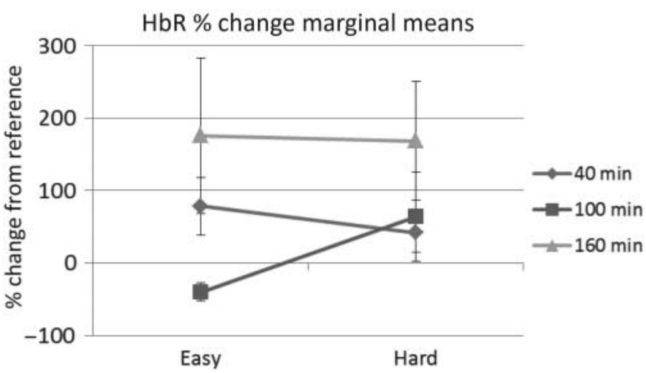

Figure 4. HbR % change shows 100-min case significantly lower than the 160-min in both easy and hard conditions, but only in easy condition for 40 min. No significant differences in HbR response for difficulty. Error bars indicate standard error.

to the average of the maximum. However, the results are expectedly not as strong.

Two-factor ANOVAs for percent change in HbO and HbR with an alpha of 0.05 found that difficulty was not a significant factor in HbO or HbR response in terms of percent changes, but that onset time was a significant factor in both HbO ($F(2, 24) = 7.641$, $P = 0.003$) and HbR ($F(2, 24) = 3.304$, $P = 0.054$), which can be seen in Figs. 3 and 4. The most striking result was that the 100-min onset time was found to have a lower response than the 40- and 160-min cases, indicating a diminished response for participants during the middle of the experiment. Figure 5 shows the data displayed in boxplot form.

3.3. Performance and workload

Track error performance for the independent variables of onset time and difficulty was analyzed using the Mann–Whitney test

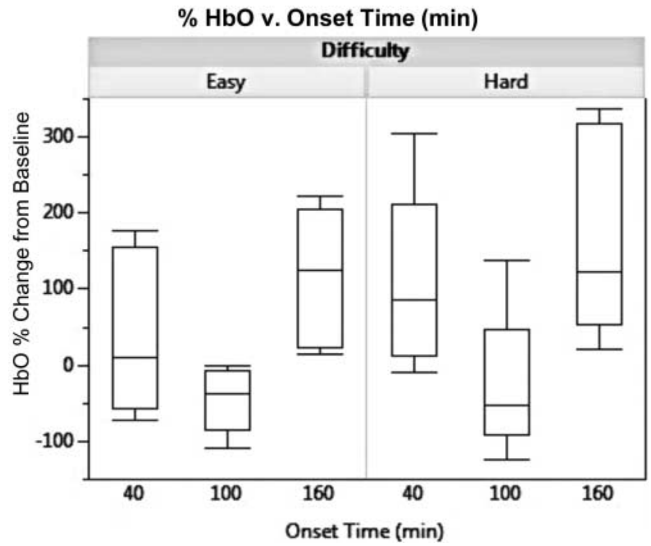

Figure 5. Box plot indicating HbO % change from reference for the six experimental conditions.

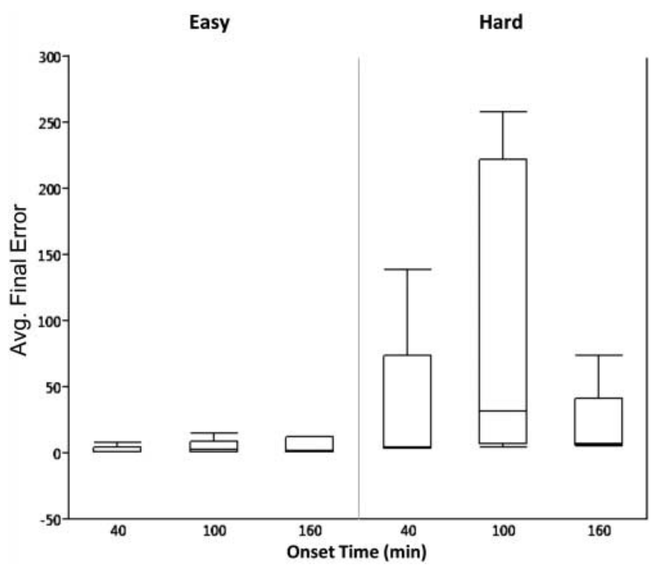

Figure 6. Box plot of average final track error score by time and difficulty. There was a significant effect for difficulty, but not onset time. The 100-min hard condition had the worst performance.

($U = 37.00$) and Wilcoxon test ($W = 157.00$). As expected, difficulty was found to be significant for final track error ($P = 0.002$), but onset time was not significant for performance ($P = 0.311$). However, Fig. 6 shows that the '100-min, 6-missile' condition was clearly linked to worse performance than all other conditions. Video analysis of the participants in this condition did not reveal any obvious anomalous behavior, such as sleeping or excessive movement.

Participant workload was measured through secondary tasking via response times to incoming messages in the text 'chat' messaging interface during the missile wave, which is both realistic to such operational settings, but also is an effective measure of spare mental capacity (Cummings and Guerlain, 2004). Because conscientiousness may be associated with participants feeling compelled to respond, a one-way ANCOVA was conducted to investigate differences in chat responses related to difficulty level and event onset, controlling for conscientiousness. The covariate, conscientiousness, was significantly related to the chat responses, $F(1, 23) = 7.049$, $P = 0.0148$. There was also a significant effect of difficulty on chat response after controlling for the effect of conscientiousness, $F(1, 23) = 4.52$, $P = 0.04$, meaning that those participants under the hard condition took significantly longer to respond to incoming messages during the wave (easy $M = 11$ s, s.d. $= 6$ s, difficult M 16 s, s.d. 9 s). Onset time was marginally significant ($F(2, 23) = 2.82$, $P = 0.08$), driven primarily by the response times for participants in the 100-min hard conditions who, on average, responded in ∼21 s while everyone in the remaining conditions took ∼11 s to respond.

4. DISCUSSION

This study has several interesting findings. First, onset time was found to have significant effects on hemodynamic response while scenario difficulty did not. This finding differs from previous studies (Bunce *et al*., 2011; Sassaroli *et al*., 2008; Solovey *et al*., 2009; Tsunashima and Yanagisawa, 2009), which suggested that fNIRS applied to the prefrontal cortex is measuring mental workload. The performance and secondary workload measures clearly show that participants struggled with the six missile condition, which required them to constantly monitor and switch the sensors across the missiles, as opposed to the three missile condition which matched the sensor resources exactly and never required any switching of resources. Thus, the fNIRS data were unable to capture the validated increase in workload, which was evident in performance and secondary workload measures. It is possible that the absence of expected differences in workload could be partially attributed to variability introduced by individual differences in participants, especially as there are five participants in each cell. However, we did not see find any obvious anomalies in the groups as shown in Table 1. This dissociation between fNIRS data and the performance and subjective workload measures indicates that further work is needed to better understand the relationship between workload, task performance and the hemodynamic response measured by fNIRS. Matthews *et al*. (2015) discuss the divergence in multiple psychophysiological measures of mental workload by exploring the sensitivity as well as intercorrelations among electrocardiogram, heart rate and heart rate variability, EEG, cerebral blood flow velocity, oxygen saturation (measured by fNIRS), eye tracking metrics, as well as NASA-TLX. They found that while some metrics were sensitive to changes in workload, the various metrics did not necessarily correspond with one another, which puts into question whether workload is the latent factor.

The unexpected finding in Figs. 3 and 4 shows that those participants in the 100-min group, under both difficulty conditions, had an unexpected net decrease in blood oxygenation in the 100 s after the onset of the critical event. Those in the 40 min onset condition were relatively fresh in the experiment and those in the 160 min onset condition knew they were only 20 min from the end of the experiment (because of the consent form). Those in the 100 min condition were just a little more than halfway through the experiment, with no immediate expectations for any change in the environment. Previous work has also found attentional inefficiencies to be highest at the relative middle of a similar long duration experiment (Hart, 2010). It is striking that 9 out of 10 participants in this condition not only did not increase their blood oxygenation, but in fact decreased which is antithetical to hypothesis that the critical event should have caused them to become more, not less, cognitively engaged. The performance results in Fig. 4 demonstrate that especially under the more difficult six missile scenario, this lack of cognitive engagement (as measured by blood oxygenation in the prefrontal cortex) led to significantly reduced performance.

The 100-min participants were at the lowest level of priming and engagement at the time of the event, and thus, had the hardest time making the mental transition from low to high task load. Thus, these results suggest that humans in such long duration, low event settings have the most difficulty transitioning from low to high workload not at the end of a shift, but rather at the point of lowest engagement, a point that often occurs somewhere in the middle.

While fNIRS and the BOLD signal are limited by the hemodynamic response rate of the brain, which can take 5–10 s following the onset of an activity, fNIRS provides a valuable and relatively non-intrusive measure of brain activity that can predict task performance. However, this time delay should be taken into account when doing such predictions or when using fNIRS data as a real-time input to an adaptive system. In addition, when moving into real-world settings, it is important to note that fNIRS is susceptible to other limitations such as major head movement, facial movement and probe movement on the forehead. Solovey *et al*. (2009) showed these limitations can be controlled in a desktop computer setting or filtered out, making fNIRS suitable for deploying in realistic settings.

5. CONCLUSIONS

In a growing number of fields, humans will face the task of monitoring a semi-autonomous system for extended periods with only occasional or rare interventions in complex, critical situations. Continuous physiological monitoring of the brain could help to both detect anomalous mental states that can degrade optimal performance during critical events, as well as generate predictions for possible degraded operators performance that could be useful in near real-time. If such reliable predictions could be achieved, then adaptive automation solutions could be implemented to prevent operators from performance degradation through some kind of active intervention.

This research explored the low-to-high workload transition problem by measuring the psychophysiological response of participants during a long duration simulated missile defense exercise. While fNIRS responses did not correlate with mental workload, the results show that hemodynamic response was diminished during the middle of a shift when engagement and priming was lowest. Future work is needed to determine how a decrease in oxygenated hemoglobin in the presence of deoxygenated hemoglobin and other demographic factors could be leveraged to potentially develop screening and/or real-time predictive monitoring tools.

ACKNOWELEDGEMENTS

Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the US Government. Rob Jacob and Dan Afergan at Tufts also assisted in this research.

FUNDING

MIT Lincoln Laboratory sponsored this work via the Missile Defense Agency under Air Force Contract #FA8721-05-C-0002.

REFERENCES

Afergan, D., Peck, E.M., Solovey, E.T., Jenkins, A., Hincks, S.W., Brown, E.T., Chang, R. and Jacob, R.J.K. (2014) Dynamic Difficulty Using Brain Metrics of Workload. Proc. ACM Conf. Human Factors in Computing Systems – CHI '14, Toronto, CA, pp. 3797–3806.

Alves, E.E. and Kelsey, C.M. (2010) Combating Vigilance Decrement in a Single-Operator Radar Platform. Ergon. Des. Quart. Hum. Factors Appl., 18, 6–9.

Ayaz, H., Cakir, M.P., Izzetoglu, K., Curtin, A., Shewokis, P.A., Bunce, S.C. and Onaral, B. (2012) Monitoring Expertise Development During Simulated UAV Piloting Tasks Using Optical Brain Imaging. IEEE Aerospace Conf., Big Sky, MT, pp. 1–11.

Bainbridge, L. (1983) Ironies of automation. Automatica, 19, 775–779.

Berka, C., Levendowski, D.J., Ramsey, C.K., Davis, G., Lumicao, M.N., Stanney, K., Reeves, L., Regli, S.H., Tremoulet, P.D. and Stibler, K. (2005) Evaluation of an EEG Workload Model in an Aegis Simulation Environment. Proc. SPIE Defense and Security Symposium, Biomonitoring for Physiological and Cognitive Performance During Military Operations, Orlando, FL. SPIE: The International Society for Optical Engineering.

Broadbent, D.E. (1958) The general nature of vigilance. Percept. Commun., 108–139.

Bunce, S.C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K. and Onaral, B. (2011) Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload Foundations of Augmented Cognition. Directing the Future of Adaptive Systems, pp. 13–22. Springer.

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D. and Thomas, R. (1998) A novel method for fast imaging of brainfunction, non-invasively, with light. Opt. Express, 2, 411–423.

Cui, X., Bray, S., Bryant, D.M., Glover, G.H. and Reiss, A.L. (2011) A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. Neuroimage, 54, 2808–2821.

Cummings, M. and Guerlain, S. (2004) Using a Chat Interface as an Embedded Secondary Tasking Tool. In: Proceedings of the 2nd Annual Human Performance, Situation Awareness, and Automation Technology Conference Daytona Beach, FL, Vol. 1, pp. 240–248.

De Joux, N., Russell, P.N. and Helton, W.S. (2013) A functional near-infrared spectroscopy study of sustained attention to local and global target features. Brain Cogn., 81, 370–375.

Derosière, G., Mandrick, K., Dray, G., Ward, T.E. and Perrey, S. (2013) NIRS-measured prefrontal cortex activity in neuroergonomics: strengths and weaknesses. Front. Hum. Neurosci., 7, 583.

Durantin, G., Gagnon, J.-F., Tremblay, S. and Dehais, F. (2014) Using near infrared spectroscopy and heart rate variability to detect mental overload. Behav. Brain Res., 259, 16–23.

Dussault, C., Jouanin, J.-C., Philippe, M. and Guezennec, C.-Y. (2005) EEG and ECG changes during simulator operation reflect mental workload and vigilance. Aviat.Space Environ. Med., 76, 344–351.

Farmer, R. and Sundberg, N.D. (1986) Boredom proneness: the development and correlates of a new scale. J. Pers. Assess., 50, 4–17.

Harrison, J., Izzetoglu, K., Ayaz, H., Willems, B., Hah, S., Woo, H., Shewokis, P.A., Bunce, S.C. and Onaral, B. (2013) Human Performance Assessment Study in Aviation Using Functional Near Infrared Spectroscopy. Foundations of Augmented Cognition, pp. 433–442. Springer.

Hart, C. (2010) Assessing the impact of low workload in supervisory control of networked unmanned vehicles. Master of Science Thesis, Massachusetts Institute of Technology, Cambridge, MA.

Hart, S.G. and Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Hum. Mental Workload, 1, 139–183.

Hirshfield, L.M., Chauncey, K., Gulotta, R., Girouard, A., Solovey, E.T., Jacob, R.J., Sassaroli, A. and Fantini, S. (2009) Combining Electroencephalograph and Functional Near Infrared Spectroscopy to Explore Users' Mental Workload. Proc. 5th Int. Conf. Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience, San Diego, CA.

Huey, B.M. and Wickens, C.D. (1993) Workload Transition: Implications for Individual and Team Performance. The National Academies Press.

Huppert, T.J., Hoge, R.D., Diamond, S.G., Franceschini, M.A. and Boas, D.A. (2006) A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. NeuroImage, 29, 368–382.

Huppert, T.J., Diamond, S.G., Franceschini, M.A. and Boas, D.A. (2009) HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. Appl. Opt., 48, D280–D298.

Izzetoglu, K., Ayaz, H., Merzagora, A., Izzetoglu, M., Shewokis, P.A., Bunce, S.C., Pourrezaei, K., Rosen, A. and Onaral, B. (2011) The evolution of field deployable fNIR Spectroscopy from bench to clinical settings. J. Innov. Opt. Health Sci., 04, 239–250.

Kahneman, D. (1973) Attention and Effort. Prentice Hall.

Kramer, A.F. (1991) Physiological Metrics of Mental Workload: A Review of Recent Progress. In Damos, D.L. (ed.), Multiple-Task Performance, pp. 279–328. Taylor & Francis, London.

León-Carrión, J. and León-Domínguez, U. (2012) Functional Near-Infrared Spectroscopy (fNIRS): Principles and Neuroscientific Applications. In Bright, P. (ed.), Neuroimaging-Methods. InTech.

Mackworth, N.H. (1948) The breakdown of vigilance during prolonged visual search. Q. J. Exp. Psychol., 1, 6–21.

Matthews, G., Reinerman-Jones, L.E., Barber, D.J. and Abich, J. (2015) The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum. Factors, 57, 125–143.

McCarthy, G., Blamire, A.M., Puce, A., Nobre, A.C., Bloch, G., Hyder, F., Goldman-Rakic, P. and Shulman, R.G. (1994) Functional Magnetic Resonance Imaging of Human Prefrontal Cortex Activation During a Spatial Working Memory Task. Proc. Natl Acad. Sci. USA, 91, 8690–8694.

McCrae, R. and Costa Jr, P. (1999) A Five-factor Theory of Personality. Handbook of Personality: Theory and Research, Vol. 2, pp. 139–153. The Guilford Press.

Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci., 24, 167–202.

Pattyn, N., Neyt, X., Henderickx, D. and Soetens, E. (2008) Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? Physiol. Behav., 93, 369–378.

Sassaroli, A., Zheng, F., Hirshfield, L., Girourd, A., Solovey, E.T., Jacob, R. and Fantini, S. (2008) Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. J. Innov. Opt. Health Sci., 01, 227–237.

Schmorrow, D. (ed.). (2005) Foundations of Augmented Cognition. Lawrence Erlbraum Associates, Mahwah, NJ.

Scholkmann, F., Klein, S.D., Gerber, U., Wolf, M. and Wolf, U. (2014) Cerebral hemodynamic and oxygenation changes induced by inner and heard speech: a study combining functional near-infrared spectroscopy and capnography. J. Biomed. Opt., 19, 017002–017002.

Schroeter, M.L., Kupka, T., Mildner, T., Uludağ, K. and von Cramon, D.Y. (2006) Investigating the post-stimulus undershoot of the BOLD signal—a simultaneous fMRI and fNIRS study. NeuroImage, 30, 349–358.

Solovey, E.T., Girouard, A., Chauncey, K., Hirshfield, L.M., Sassaroli, A., Zheng, F., Fantini, S. and Jacob, R.J.K. (2009) Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines, Proc. ACM Symp. User Interface Software and Technology UIST '09, Victoria, BC, pp. 157–166.

Solovey, E., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S. and Jacob, R. (2012) Brainput: Enhancing Interactive Systems with Streaming FNIRS Brain Input. Proc. ACM Conf. Human Factors in Computing Systems CHI'12, Austin, TX.

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S. and Obrig, H. (2006) Illuminating the BOLD signal: combined fMRI–fNIRS studies. Magn. Reson. Imaging, 24, 495–505.

Strangman, G., Culver, J.P., Thompson, J.H. and Boas, D.A. (2002) A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. NeuroImage, 17, 719–731.

Tsujimoto, S., Yamamoto, T., Kawaguchi, H., Koizumi, H. and Sawaguchi, T. (2004) Prefrontal cortical activation associated with working memory in adults and preschool children: an event-related optical topography study. Cereb. Cortex, 14, 703–712.

Tsunashima, H. and Yanagisawa, K. (2009) Measurement of brain function of car driver using functional near-Infrared spectroscopy (fNIRS). Comput. Intell. Neurosci., **2009**, DOI: 10.1155/2009/164958.

Veltman, J. and Gaillard, A. (1996) Physiological indices of workload in a simulated flight task. Biol. Psychol., 42, 323–342.

Warm, J.S., Parasuraman, R. and Matthews, G. (2008) Vigilance requires hard mental work and is stressful. Hum. Factors: J. Hum. Factors Ergon. Soc., 50, 433–441. Human Factors and Ergonomics Society.

Warm, J.S., Matthews, G. and Parasuraman, R. (2009) Cerebral hemodynamics and vigilance performance. Mil. Psychol., 21, S75–S100.

Warm, J.S., Tripp, L.D., Matthews, G. and Helton, W.S. (2012) Cerebral Hemodynamic Indices of Operator Fatigue in Vigilance. Handbook of Operator Fatigue, pp. 197–207.

Wickens, C.D. (1984) Processing Resources in Attention. In Parasuraman, R. and Davies, D.R. (eds), Varieties of Attention, pp. 63–102. Academic Press, Orlando, FL.

Wilson, G. (2002) An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int. J. Aviat. Psychol., 12, 3–18.

Wolf, M., Ferrari, M. and Quaresima, V. (2007) Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. J. Biomed. Opt., 12.